Target Article

Mundari: The myth of a language without word classes

NICHOLAS EVANS and TOSHIKI OSADA

Abstract

Mundari, an Austroasiatic language of India (Munda family), has often been cited as an example of a language without word classes, where a single word can function as noun, verb, adjective, etc. according to the context. These claims, originating in a 1903 grammar by the missionary John Hoffmann, have recently been repeated uncritically by a number of typologists. In this article we review the evidence for word class fluidity, on the basis of a careful analysis of Hoffmann's corpus as well as substantial new data, including a large lexical sample at two levels of detail. We argue that in fact Mundari does have clearly definable word classes, with distinct open classes of verb and noun, in addition to a closed adjective class, though there are productive possibilities for using all as predicates. Along the way, we elaborate a series of criteria that would need to be met before any language could seriously be claimed to lack a noun-verb distinction: most importantly strict compositionality, bidirectional flexibility, and exhaustiveness through the lexicon.

Keywords: *conversion, derivation, kinship terms, Mundari, noun, omnipredicative, precategoriality, proper name, verb, word classes, word formation*

1. Introduction

The possibility that there exist languages lacking a noun-verb distinction is not only the most extreme challenge to universalizing theories of word classes, but it also raises profound philosophical issues about whether all humans find the cognitive distinction between objects and events to be self-evident (cf. Whorf 1956 [1940]). The claim that there are languages lacking a distinct open class of

Linguistic Typology 9 (2005), 351–390 1430–0532/2005/009-0351

adjectives is now well-established (Dixon 1977,[1] Schachter 1985, Croft 2003). But the search for the more extreme case of a language with just a single open word class of predicates has not yielded an uncontroversial example, to the point where the second edition of Croft's influential textbook on typology and universals states confidently that "one of the few unrestricted universals is that all languages have nouns and verbs" (Croft 2003: 183).

Yet the phenomenon of fluid word class membership – of languages claimed to lack a noun-verb distinction entirely, or to have only a very weak noun-verb distinction – has recently experienced renewed attention. The longstanding debate over whether the noun-verb distinction exists in certain languages of the Austronesian family, and in the Pacific coast of the northwestern U.S. and southwestern Canada, has recently been resurrected with detailed new analyses by Himmelmann (1991, 2004a, b) and Gil (1995) for Tagalog, by Gil (1994, 2001) for Riau Indonesian, by Mosel & Hovdhaugen (1992) for Samoan, by Broschart (1997) for Tongan,[2] and by Kuipers (1968), Jacobsen (1979), Kinkade (1983), Jelinek & Demers (1994), Jelinek (1995), and Demirdache & Matthewson (1995), among others, for Salish and other languages of the Pacific Northwest. These studies have brought new sophistication to our understanding of the often very delicate issues of analysis required to decide whether these languages merely have a more subtle difference between nouns and verbs, or have a morphological distinction but no syntactic distinction, or have a clear distinction but highly productive rules of zero conversion, or simply have a single major word class of predicates.

A typology of ways in which languages may blur the distinction between the major word classes of nouns and verbs will be presented in Section 1.2, and – because debate in this area so often employs incompatible argumentation and terms – in Section 2 we set out some general criteria that need to be met before any claim about lack of word class distinctions can be deemed proven. We then move, in Section 3, to the main part of our paper: a re-examination of the status of the noun-verb distinction in Mundari, another language for which there is a long history of arguments against the existence of a noun-verb distinction. Mundari belongs to the Munda branch of Austroasiatic (Figure 1) spoken in northern India by around three quarters of a million people; see Osada (1992) for details.

---

1. A recent collection edited by Dixon & Aikhenvald (2004) retreats from this position, arguing that adjectives are, after all, a universal class, though not necessarily an open one. Some of the contributors in that volume, however, do not take this position (see, e.g., Enfield's chapter on Lao), regarding adjectives as simply a subclass of stative verbs. Since the status of the adjective class is not the focus of this paper we do not pursue this issue further here.
2. Though for another Oceanic language, Fijian, the more detailed argumentation in Dixon (1988) makes a clear case that an earlier monocategorialist analysis by Milner (1972) is unsustainable.

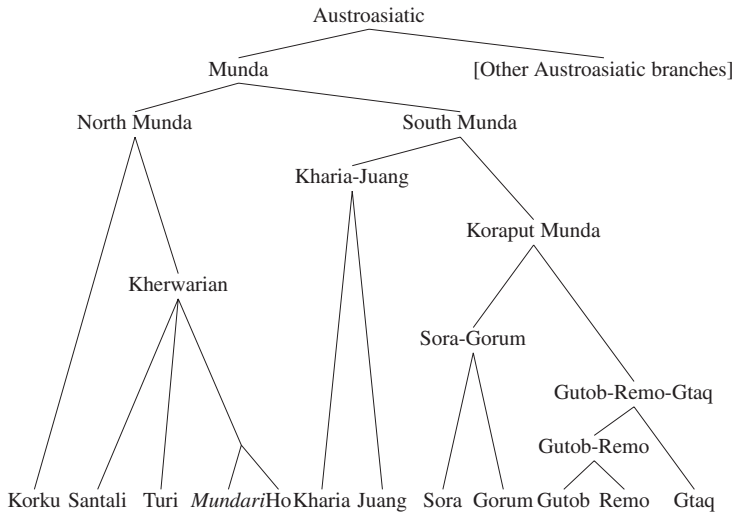

Figure 1. *The Munda languages*

A re-examination of the Mundari case is long overdue, because the renewal of interest in fluid word class membership has not, so far, brought any new data or argumentation to claims that go back to 1903. In fact, several prominent typologists have recently given Mundari as an example of a language with a very fluid word class system: Hengeveld (1992a, b), Stassen (1997), Wetzer (1996), and Rijkhoff (2002, 2003), among many others (see below for further examples), cite Hoffmann's (1903) Mundari grammar, without any evident reservations, as exemplifying a language where it is impossible to assign words to clearly defined parts of speech.

The following quote from Hoffmann (1903: xxi) sums up his original position:

> Thus the same unchanged form is at the same time a Conjunction, an Adjective, a Pronoun, an Adverb, a Verb and a Noun, or, to speak more precisely, it may become a Conjunction, an Adjective, etc., etc.; but by itself alone it is none of them. It is simply a vague elastic word, capable of signifying, in a vague manner, several distinct concepts, i.e. of assuming a variety of functions.

Many authors have repeated similar positions with regard to Mundari and other Munda languages.[3] For example, Pinnow (1966: 101) states:

3. See also Sinha (1975). By contrast, Neukom's (2001) recent grammar of Santali effectively recognizes the existence of a two-way major word class distinction, though so far his analysis

> Theoretically any word for any concept, i.e., all words, can function as a verb base. Thus we may not speak of a verb in the Indo-European sense. This fact was recognized at an early date and is now generally known. [. . .] This phenomenon undoubtedly goes back to very ancient times and can probably be accepted as Proto-Munda.

Hengeveld (1992: 47) characterizes Mundari as a "flexible" language, with a single V/N/A/Adv category; Hoffmann (1903) is the only source cited. And Bhat (1997), in a recent restatement of Hoffmann's position (again based just on data in his 1903 grammar), concludes for Munda (1997: 249) that "the noun-verb distinction can only be viewed as a functional one in this language; it has not been lexicalized as in English and other familiar languages". Similar statements by the same author can be found in Bhat (1994, 2000: 56–57).

To get a feeling for the reasoning behind these views, consider the following four Mundari sentences, which are typical of those used in the above discussion. In (1) *buru* is used as an argument, with the meaning 'mountain', while in (2) it is used as a two-place predicate with the meaning 'heap up'. To illustrate the other direction of deployment, in (3) the word *jom* is used as a two-place predicate with the meaning 'eat', while in (4) it is used as an argument with the meaning 'food'. These sentences, incidentally, illustrate the main features of Mundari grammatical organization: there is a clause-final predicate with a complex series of affixes for aspect, transitivity, and mood (as well as a number of other categories not shown here), preceded by argument NPs, usually in the order SOV when full NPs are involved (5), though the subject is typically omitted.

(1) *buru=ko* *bai-ke-d-a.*[4]
mountain=3pl.S make-COMPL-TR-INDIC
'They made the mountain.'

---

does not appear to have been absorbed into the general typological literature. Neukom argues that even though most words can be used in predicate position, "there is a group of lexemes which cannot be determined by demonstratives", which he terms verbs. He proposes (2001: 17) that verbs constitute a relatively large group, around a third of all lexemes, citing twelve forms as examples – *əgu* 'bring', *əiku* 'feel', *bəgi* 'leave', *bɔlɔ* 'enter', *caŋke* 'be ravenous', *cet'* 'learn, teach', *dɔhɔ* 'put', *gitic'* 'lie, lay', *lɔ* 'burn', *orom* 'find out', *rəput'* 'break', *sɛn* 'go'. He also invokes the equivalent of our "bidirectional" criterion, pointing out (2001: 13) that "[i]f we assume that there is only one lexeme class in Santali, every lexeme should appear in both positions: in argument and in predicate position". A difference between Neukom's analysis of Santali and our analysis of Mundari is that although he does recognize a class of verbs he has no class of nouns; rather, he has a "lexeme or lexeme combination" class which may behave either as argument phrases or as predicates; it is a little unclear where this leaves verbs, which are certainly lexemes, so that a more accurate term may have been "flexible lexemes", as opposed to verbs. In any case, this analytic decision reflects in part the greater freedom to use any lexeme as predicate in Santali, though it may also reflect differences in argumentation that we do not pursue here.

(2) *saan=ko* *buru-ke-d-a.*
firewood=3pl.S "mountain"-COMPL-TR-INDIC
'They heaped up the firewood.'

(3) *maNDi=ko* *jom-ke-d-a.*
food=3pl.S eat-COMPL-TR-INDIC
'They ate the food.'

(4) *jom=ko* *nam-ke-d-a.*
"eat"=3pl.S get-COMPL-TR-INDIC
'They got the food.'

Subjects are cross-referenced by enclitics, normally placed in preverbal position when there is material to host them, as in (1)–(5), but these come after the verb if there is no preverbal material available as host (6a), and there is an increasing trend for younger speakers to place them after the verb even when other material is present (6b).

(5) *seta-king* *pusi-ko=king* *hua-ke-d-ko-a.*
dog-DU cat-PL=3du.S bite-COMPL-TR-3pl.O-INDIC
'The two dogs bit the cats.'

(6) a. *hua-ke-d-ko-a=king.*
bite-COMPL-TR-3pl.O-INDIC=3du.S
'The two of them bit them.' [requires context to establish pronominal reference]
b. *seta-king* *pusi-ko* *hua-ke-d-ko-a=king.*
dog-DU cat-PL bite-COMPL-TR-3pl.O-INDIC=3du.S
'The two dogs bit the cats.'

Though the elements following the verb in examples like these have generally been regarded as suffixes, they are in fact less tightly bound to the root, phonologically, than other suffixes are. Thus monosyllabic words of form CVʔ add a (non-phonemic) echo vowel when unsuffixed (7), but when a "close" suffix is present, such as the passive suffix *-oʔ*, there is no echo vowel and the glottal stop is replaced by /g/ (8). Before predicate inflections the echoic vowel is found and the glottal stop remains unchanged (8b, c). Disyllabic glottal-final

4. A practical orthography is employed here: ng = ŋ, ñ = ɲ, q = ʔ, retroflexion shown by capitalization, e.g., T = ʈ but t = t. Since we employ capital letters to indicate retroflexion, we refrain from capitalizing the initial letters of sentences in this transcription. Phonemic vowel length is shown by doubling the vowel. The "checked" realization of word-final stops, phonetically realized as pre-glottalization or a subsequent nasal release, is phonemically predictable and not shown here, e.g., we write *sab* 'catch' for [saʔp$^{(m)}$]. In Hoffmann's works this word would be written with a wedge under the *b*: [sa̰b].

roots, however, which lack the echo vowel in citation form, convert the glottal stop into /g/, like with passive suffixes (9), so the difference in tightness of bonding is only evident with monosyllables.

(7)
a.	/daʔ/	[daʔa]	‘water’
b.	/kuʔ/	[kuʔu]	‘cough’
c.	/maʔ/	[maʔa]	‘cut with axe’

(8)
a. *kug-oq-ta-n*[5]*-a=eq*
cough-PASS-PROGR.OR-INTR-INDIC=3sg.S
‘(S)he coughed (involuntarily, e.g., through food getting stuck in her throat).’

b. [*kuʔu*]*-ta-n-a=eq*
cough-PROGR.OR-INTR-INDIC=3sg.S
‘(S)he coughed.’

c. *daru=ñ* [*maʔa*]*-ke-d-a*
tree=1sg.S cut-COMPL-TR-INDIC
‘I cut the tree.’

(9)
a. *setaq* [*setaʔ*] ‘morning’

b. *setag-aka-n-a.*
morning-INIT.PROG-INTR-INDIC
‘It grew morning, it dawned.’

Returning to examples (1)–(4), the argument advanced in Hoffmann’s 1903 grammar, and repeated in the above secondary sources, is that the possibility of using words like *buru* or *jom* in ways that correspond to either nouns or verbs in English shows that they lack any inherent word class, and that such word classes as noun, verb, or adjective are Eurocentric impositions that cannot justifiably set up for the language.

Now it is remarkable that, despite the great typological importance of the Mundari case, the many recent mentions in the typological literature all rely on Hoffmann’s grammar – and that does not even represent the mature view of the author they cite, let alone more recent treatments (e.g., Osada 1992). Hoffmann was a missionary who lived among the Mundari for twenty-two years, and then spent a further thirteen years in Europe revising his massive *Encyclopaedia Mundarica* for publication after being evacuated from India during the First World War, as a German whose nationality was deemed to make him an unreliable British patriot.[6] His grammar of Mundari – the work cited by

5. With intransitive verbs, the passive produces a non-volitional reading.
6. After leaving India Hoffmann was aided by various collaborators in India, who continued the work in the decades following his death in 1928; see Ponette (1990).

Hengeveld, Stassen, Bhat, and others – was thus a relatively immature work, written in only 1903, and it takes the extreme monocategorialist (or better, precategorialist – see below) position cited above. But by the time he came to work on the sixteen-volume, 4889-page *Encyclopaedia Mundarica*, with decades of further work on the language behind him, Hoffmann had retreated from his earlier and more radical position: he lists words with word class labels, and in places states explicitly that certain words must be used in particular functions. We will return to this in detail in Section 3.4.

In this article we will draw extensively on data from *Encyclopaedia Mundarica*, as well as fieldwork by Osada carried out in the Ranchi region since 1984, supplemented by focussed checking carried out by both authors with Maki Purti, who speaks Hindi, English, and Japanese in addition to Mundari. This further checking focusses on the two major (possible) word classes of noun and verb. Both sets of additional data necessitate a re-evaluation of Hoffmann's 1903 claims.

The new data we examine shows that Mundari, like all the Munda languages, makes wide use of zero conversion, resulting in frequent heterosemy, i.e., the use of identical forms with different combinatorics and different meanings (Lichtenberk 1991), and this is as true of variant combinatorics within word classes as it is across them. For example, placing basically intransitive verb roots within a transitive predicate frame is a common way of forming causatives. The use of the same affixal forms across word classes is also widespread, such as the use of the same infix ⟨*pV*⟩ for reciprocals (with verbs) and intensification (with adjectives), or the same bound pronominal forms for possession (when suffixed to nouns), to mark subject agreement (when encliticized to the last preverbal constituent), object agreement (when suffixed after the transitivity marker on the verb), or indirect object agreement (when suffixed before the transitivity marker).

A tempting analysis, upon initial inspection of facts of this type, is to see Mundari lexemes as signs whose signifiers are fixed, but whose combinatorics are unspecified, and whose signifieds are only partially specified – in other words, to treat lexemes as precategorial, with some underspecified meaning present in the lexicon, but with the balance supplied from whichever syntactic frame they find themselves in. Hoffmann's term "a vague, elastic word, capable of signifying, in a vague manner, several distinct concepts" suggests he saw the system in this way, and it is also implicit in Bodding's (1929) grammar of the closely-related Santali language, and in Bhat's (1994, 1997, 2000) restatement of the Mundari word class problem, on the basis of Hoffmann's data.[7] We shall refer to this class of analyses as precategorial.

---

7. Additionally, this is essentially the analysis proposed by Schiller (1989, 1992) for another Austroasiatic language, Khmer, within an autolexical framework.

This may look appealing if we confine ourselves to a few choice cases and do not require that regular and predictable semantic increments accompany the specification of word class. But once we extend our investigation to a wider set of lexical items, and take seriously the requirement that the semantic effects of category specification be fully compositional, and equivalent across semantically comparable items, the most plausible analysis of Mundari is as a language with clear noun, verb, and adjective classes, but a great deal of zero conversion, often lexically idiosyncratic. We conclude that, while languages lacking a noun-verb distinction may well exist – we would love to find one! – Mundari is not a plausible candidate. Along the way, we hope to lay out some explicit rules of argumentation that any seeker of such a language would need to satisfy before they can rest their case.

2. Ways a language could lack a noun-verb distinction

Because the descriptive and typological literature conflates a number of phenomena under the general rubric of word class fluidity, with confusing consequences for what gets cited as evidence against there being basic word class distinctions, in this section we outline four ways that a language can lack – or appear to lack – a noun-verb distinction, moving from the strongest to the weakest case.

Note that these possibilities are orthogonal to the question of which domain of grammar the evidence is drawn from, and that for any one of our four types one needs, in principle, to distinguish morphological from syntactic evidence[8] and to leave open the possibility that word classes distinguishable by morphological criteria could be indistinguishable by syntactic criteria (cf. Evans 2000a). Obviously, in the strongest, ideal case of a language lacking a noun-verb distinction, words would have the same behaviour both morphologically and syntactically, but it is helpful to be able to use languages where the relevant criteria are only met at the syntactic level.

8. An anonymous *LT* reviewer objects to our separation of morphological from syntactic evidence, on the grounds that "most grammatical phenomena examined in typological studies are manifested either morphologically or syntactically or a combination of the two", and that "grammaticalization theory also argues against any sharp division". We stand by our separation. Even though syntax can evolve into morphology over time, there are clear distinctions between the two in any synchronic state of the language (leaving aside the possible existence of some awkward boundary cases in particular languages), so that the two provide quite different synchronic diagnostics for lexical categories.

2.1. *Omnipredicative languages*

We adapt this helpful term from Launey (1994) to describe the situation where all major word classes are able to function directly as predicates without derivation, and with no change of meaning. Note that the description of Classical Aztec for which Launey introduces this term readily recognizes the existence of distinctive word classes, on the basis of the clearly distinct morphological possibilities of nouns and verbs, even though both can function directly as predicates. On our more general use of his term, an omnipredicative language would be one in which all major-class lexical items belong to a single word class of "predicates", with no morphological differences such as are found in Classical Aztec.

The clearest illustration of how such a language would work does not come from a natural language, but from an artificial language, Predicate Calculus, in which the exponents of 'run', 'big', and 'man' are all simply one-place predicates with identical syntactic possibilities: RUN (*x*), BIG (*x*), and MAN (*x*) to express the English propositions '*x* runs', '*x* is big', and '*x* is a man'. (Obviously we need to supplement the predicate calculus with a tense logic, as well as appropriate devices for representing definiteness, before the English propositions can be said to be faithfully represented.) Predications can then be nested in appropriate ways to construct the representation of what, in English, would employ a clause with a verbal predicate, and a subject noun phrase made up of a noun and an adjective:[9]

(10) Run (x: (Man (x) & Big (x)))
'(The) big man runs.'

On Launey's description of Classical Nahuatl this requirement is met at the syntactic but not at the morphological level.

Morphologically, nouns are clearly distinguished from verbs by the availability of a series of tense-aspect-mood suffixes on verbs only, as well as a range of other morphological possibilities not shown here (most importantly pronominal object agreement, applicatives, reflexive-reciprocal marking, causatives, noun incorporation) and of the absolutive suffix for (non-possessed) nouns only (11a–c).

---

9. It is a frequent assumption that the two main distinctive functions of adjectives – to restrict reference, when used attributively, and to achieve copredication with attendant time-boundedness under clausal tense, when used as secondary predicates – can be assimilated to the basic functions of predication and reference (cf. Thompson 1988, Croft 2003). There are of course problems with this assumption, for example explaining the syncategorematicity effects found with many adjectives but not typically with members of other word classes; but we do not pursue these here.

(11) a. *Ø-chōca in Ø-piltōn-tli*
3SG-cry DEM 3SG.S-child-ABS
'The child cries.' (Launey 1994: 29)[10]
b. *Ø-chōca-z in Ø-piltōn-tli*
3SG-cry-FUT DEM 3SG.S-child-ABS
'The child will cry.' (Launey 1994: 29)
c. *Ø-chōca-ya in Ø-piltōn-tli*
3SG-cry-P.IMPF DEM 3SG.S-child-ABS
'The child was crying.' (Launey 1994: 29)

Syntactically, however, both nouns and verbs have equivalent possibilities for being employed in predicate or argument slots: both take (identical) person-number prefixes for their subjects (including 3rd singular zero), both may serve equally as arguments or predicates, and both nouns and verbs must equally be preceded by the referentializing demonstrative *in* when in argument roles.

(12) a. *ni-chōca*
1SG.S-cry
'I cry.' (Launey 1994: 42)
b. *ti-chōca*
2SG.S-cry
'You cry.' (Launey 1994: 42)
c. *Ø-chōca*
3SG.S-cry
'He cries.' (Launey 1994: 42)

(13) a. *Ø-tzat'tzi in Ø-konē-tl*
3SG.S-shout DEM 3SG.S-child-ABS
'The baby shouts.' (i.e., he shouts, the one who is a baby)
(Launey 2002: 115)
b. *Ø-konē-tl in Ø-tzat'tzi*
3SG.S-child-ABS DEM 3SG.S-shout
'It is a baby who is shouting.' (Launey 2002: 115)

(14) a. *ni-c-yōllālia in Ø-chōca*
1SG.O-3SG.O-console DEM 3SG.S-cry
'I console the one who cries.' (Launey 1994: 59)
b. *Ø-tlaìiyōhuia in Ø-chōca*
3SG-suffer DEM 3SG-cry
'He who cries suffers.' (Launey 1994: 59)

10. We have added interlinear glosses to Launey's examples, including zero 3rd person singular subject prefixes, in line with other sources on Classical Aztec such as Andrews (1975), and translated his French translations into English.

Another omnipredicative-style description of a language is Jelinek's (1995) analysis of Straits Salish, in which all major-class lexical items are said to simply function as predicates, of the type 'run', 'be big', 'be a man', and so forth. They are then slotted into various roles in the clause, such as argument ('the one such that they run'), predicate ('run(s)'), and modifier ('the one running'), according to the syntactic slots they are placed in. The single open syntactic class of predicate includes words for entities (15), qualities (16), and events (17). When used directly as predicates, all appear in clause-initial position, followed by subject and/or object clitics (which may be zero in the case of 3rd person singular).

(15) *swiʔqoaɬ-lə=sx*^{w}
be.young.man-PERF=2SG
'You were a young man.'

(16) *ʔəy'=Ø*
good=3
'He is/was good.'

(17) *yeʔ-ə-sə=sx*^{w}
go-QUESTION-FUTURE=2SG
'Will you go?'

When used as arguments, all are effectively converted into relative clauses through the use of a determiner, which must be employed whether the predicate word refers to an entity (18), an event (19), or even a proper name (20).

(18) *ŋa-t=Ø=sən* *cə* *sčeenəx*^{w}
eat-TR=3.ABS=1.NOM DET fish
'I ate the/a/some) fish.'

(19) *cəsəʔ* *cə* *t'iləm*
be.two DET sing
'They are two, the ones who sang.'

(20) *niʔ* *sə* *Eloise*
exist DET.FEM be.Eloise
'Eloise was born.' (lit. 'Exist(ed) the being-Eloise one')

For a language to be established as omnipredicative, then, all words not belonging to minor word classes (such as determiners, grammatical particles, and perhaps some closed adverbial classes), should be able to function directly as predicates, and should have equal potential to form referring expressions through relativization or at least the addition of some sort of determiner. A semantic corollary is (i) that the base meaning of words not denoting actions

should take the form 'be [X], be [Y], be [Z]' (where X, Y here represent meanings typically expressed by common or proper nouns, or adjectives, in languages like English), and (ii) that using words from the predicate class as arguments should produce no further semantic increment than that accompanying relativization in English, i.e., '(the) one that (is) P'.

Our brief discussion of the Nahuatl and Salish examples is not supposed to indicate that, in either case, the argument for omnipredicativity has been fully established, since in both cases the respective authors admit that there are morphological differences, and in neither case do they carry out a comprehensive survey of the open-class lexicon to show that their analysis is completely productive, rather than being limited to a few well-behaved lexemes. Rather, it is supposed to show what an omnipredicative language would look like in principle.

2.2. *Precategorial languages*

The word "precategorial" has been used in a variety of ways in the literature, often rather loosely.[11] In this article we will not try and adjudicate between these various uses or analyse all the cases it has been applied to. Rather, for the purpose of illustrating a particular possible organization of word class systems, we will restrict it to the case where – as in omnipredicative languages – open-class lexemes can occur in any syntactic position. However, in precategorial languages it is not possible to state a predicate-type meaning for the lexeme

11. See, for example, Foley (1998). The closest he comes to a definitional-like statement is on p. 24: "Tagalog roots are basically precategorial, neither noun nor verb". Though this characterization is logically compatible with roots being simply predicates, earlier on the same page he suggests that, in Tagalog, "roots like *bigay* 'give', *halu* 'stir', *bili* 'buy' etc. do not entail argument structure at all, merely some generalized conceptual structure paraphrasable as 'giving by X of Y to Z' or 'stirring by A or B into C at D' … True argument structure as we understand it crosslinguistically would only be introduced when the roots are derived with the voice markers". This formulation suggests that the most important aspect of precategoriality, in Foley's view, is that it precedes the association of thematic roles with argument structure. Note that Kroeger (1998) argues against Foley's analysis of Tagalog precategoriality, on the basis of detailed and exhaustive data from Tagalog and Kimaragang (another Philippine-type language).
An attempt at restoring clarity to the notion is made in Himmelmann (2004a: 129), who (returning to the original usage of Verhaar 1984) suggests restricting it to "precategorial bound roots, i.e. lexical bases which do not occur without further affixation or outside a compound in any syntactic function and from which items belonging to different morphological or syntactic categories (nouns and verbs, for example) can be derived, without there being clear evidence that one of the possible derivations from a given root is more basic than the other one". As Himmelmann's wording makes clear, precategoriality would then be a feature of roots rather than lexemes, and is compatible with the existence of syntactic categorial distinctions between nouns and verbs.

directly; rather there is an increment that is made, according to the functional position it is plugged into.

Though he does not use the term "precategorial", Sasse's (1993) account of word classes in Salishan languages and Tagalog, extending Himmelmann's (1991) analysis of Tagalog, nicely captures what precategorial might be taken to mean in languages

> that do not preclassify lexemes for predicative and referential use, but mark the difference syntactically by establishing a predicative relation which resembles that of nominal (copula) sentences, and establishing a referential slot by using a special article-like referential marker. [...] Both languages are thus able to escape the formation of lexical categories by using these neutral expressions now predicatively, now referentially, just as we use nouns now as arguments, now as predicate nouns in nominal sentences. (Sasse 1993: 655)

Whereas the omnipredicative position discussed in Section 2.1 sees all members of the major class as basically predicates, the view just cited differs subtly by not pre-assigning the members of the one major class to a predicate role: rather, "predicativity is not inherent in the so-called 'verbs', but established syntactically by the juxtapositive linking up with a predication base ('subject')" (Sasse 1993: 661).

However, other descriptions of languages that avoid assigning inherent word class categories to lexical items contain at least some examples where the semantic distinction between nominal and verbal uses exceeds the minimal difference suggested by Sasse's treatment. An example of this analysis is Mosel & Hovdhaugen's (1992: 76) grammar of Samoan:

> Many, perhaps the majority of, roots can be found in the function of verb phrase and noun phrase nuclei and are, accordingly, classified as nouns and verbs [...] This does not mean that a noun can be used as a verb or a verb as a noun or that we have two homophonous words [...] Rather, it means that in Samoan the categorization of words into nouns and verbs is not given a priori in the lexicon.

Now while the discussion after this quote contains examples where the contrasted uses of a single lexeme only contrast on the predicate vs. referential dimension, e.g., *E uō Tanielu ma Ionatana* 'Daniel and Jonathan are FRIENDS' vs. *E alofa Tanielu i lana uō* 'Daniel loves his FRIEND' (1992: 77), it also includes examples where the semantic difference is rather greater, such as the 'thief' and 'steal' meanings of *gaoi* (1992: 77), or the 'fish-rich, fishy, successful in yielding fish' meaning found with *ia* 'fish' when used in predicate position combined with an intensifier (1992: 78). If the term "precategorial" is to be applied to a situation like this, it can only work by arguing that the meanings of the lexical items are much vaguer before they are plugged into

a syntactic context, rather as a Semitic root like *ktb* can only be given a very abstract meaning outside of a particular *binyan*.

Though we have discussed "precategorial" languages as a separate type to "omnipredicative" because of the frequent use of the term "precategoriality" in the recent literature, we have strong reservations about the provability of such claims, since a convincing model would need an explicit statement of what lexemes, or roots, would mean in such a language, but precategorialist treatments typically state that lexeme meanings are ineffable, outside their particular use in predicate or argument slots. However, since our arguments about Mundari do not depend on this point, we do not go into this critique further here. Where we want to be non-committal between omnipredicative and precategorial positions, we shall use the term "monocategorial" as a neutral term.

2.3. *Broschartian languages*

Broschart's (1997) analysis of Tongan[12] illustrates another analytic approach to languages in which the semantic result of placing lexemes in referring or predicating environments depends not on a high-level word class category like noun or verb, but rather is sensitive to much more specific semantic categories, each characterized by their own particular pattern of semantic incrementation.

For Broschart, the main syntactic distinction is between "type" and "token" expressions, with "token" expressions co-occurring with articles, while "type" expressions co-occur with TAM particles; there is an obvious parallel to more standard analyses in terms of reference and predication respectively. A number of lexical classes differ in the way their semantics is augmented when they are plugged into these syntactic environments and interact with other productive morphological formatives, but these categories reflect semantic groupings like 'action', 'task', 'personal relation', 'nationality', etc. rather than the more general groupings ('entity', 'quality', 'state/event') that traditional linguistics aligns with the word classes of noun, adjective, and verb.

Though Broschart's analysis has the appeal of not postulating higher categories (noun, verb) that – at least according to his analysis – do no work for the description of the language, and of accounting for apparent regularities in the semantic relationship between a given lexeme's differing interpretations in "type" and "token" uses, it comes at a cost: the semantic differences are too great to be attributed simply to the difference between predication and reference. For example, 'task' words with meanings like 'king' mean 'act as king' when used as predicates, 'language' words like 'Tongan' mean 'speak Tongan',

12. We cite this paper to illustrate a type of approach, rather than to endorse his analysis of Tongan word classes. See Churchward (1953) for a grammar of Tongan that gives full criteria for recognizing three major classes of noun, verb, and adjective.

'tool' words like 'hoe' mean 'use a hoe', and so forth. At best, then, the appeal of Broschart's position is to offer regular principles for polysemous extensions, based on coherent lexical groupings and what are presumably highly efficient and recurrent pairings of increments with lexical groupings (e.g., 'speak' with language names, 'use' with tools, and so forth), but because of the great range and complexity of the added semantic components it cannot get them to fall out of the frames that the lexemes are used in.

2.4. *Rampant zero conversion languages*

Moving yet further into the realm of lexical idiosyncrasy, in rampant zero conversion languages the vast majority of lexical items of a given form may appear in both predicating and referring syntactic environments with no formal signalling of conversion, but unlike in a Broschartian language, the semantic effects of syntactic environment are far less predictable. In such languages – English being a reasonable though not completely thorough-going example – it is traditional to recognize distinct word classes of noun, verb, adjective, etc., though some analysts have been reluctant to take this approach to Classical Chinese (see Norman 1988 and Evans 2000a: 724–725 for critiques). English may be used to illustrate: taking the four artefact terms *shovel*, *cup*, *can,* and *spearhead* and examining their meanings when zero-converted to transitive verbs: the first gives a 'use as instrument' meaning ('shift (coal, etc.) with or as with'),[13] the second a 'form into shape' meaning ('form into the shape of a cup (e.g., hands)')[14], the third a 'place in' meaning, and the fourth 'act like a' meaning (*the fourth battalion spearheaded the attack*). Though many regularities can still be discerned in the semantic increments accompanying zero conversion, they are too diverse and chaotic to make it worthwhile to set up single precategorial lexical meanings from which the actual meanings when used in particular syntactic contexts can be derived.[15]

We have outlined these four somewhat idealized types, which are effectively points along a continuum, in order to furnish a typological framework against which claims of languages lacking a noun-verb distinction can be placed. It is particularly noteworthy that properly assessing a language's status cannot be done with just one or two lexemes since, with the right single example, a language of any one of these four types can be made to look like an omni-

13. This and other English definitions are from the Oxford English Dictionary.
14. Though the OED lists the 'use as instrument' sense 'bleed (person) by means of a [cup]ping glass'.
15. For a fine survey of how many unpredictable, context-particular factors may contribute to the reading of zero-converted terms in English conversation (many of which may then go on to become conventionalized) see Clark & Clark (1979).

predicative language. A broad sampling of lexical items across a range of semantic categories must be made before any serious conclusions can be reached.

3. Three criteria for establishing lack of word class distinctions

Since it is only monocategorial languages, from among the types above, that can truly be claimed to lack a noun-verb distinction, we now proceed to formulate three criteria which must be satisfied before a language can be claimed to be monocategorial, illustrating with clear cases from other languages where possible, and then examining the Mundari facts.[16] Summarizing these criteria briefly, the (putative) merged classes should be distributionally equivalent (i.e., members of both classes should have equivalent combinatorics) (Section 3.1), and the semantic results of using a member of one (putative) class in a constructional slot prototypically associated with the other (putative) class should be derivable through strict compositional principles (Section 3.2). A corollary of Section 3.1, which we will discuss after Section 3.2 for expository purposes, is that the effects should be bidirectional (Section 3.3), i.e., members of X should be deployable in the environments associated with Y, and members of Y should be deployable in the environments associated with X. Finally, the preceding criteria should be exhaustive across the lexicon (Section 3.4), i.e., the same tests should yield the same results for all lexemes in the putative class, not just for a few well-chosen ones. We will see that applying these three criteria decisively demonstrates that Mundari is NOT a monocategorial language.

3.1. *Equivalent combinatorics*

This is the obvious starting point: MEMBERS OF WHAT ARE CLAIMED TO BE MERGED CLASSES SHOULD HAVE IDENTICAL DISTRIBUTIONS IN TERMS OF BOTH MORPHOLOGICAL AND SYNTACTIC CATEGORIES. Note that this has to hold for ALL combinatorics available to a function.[17] Considering our Mundari examples (1)–(4), the claim here would be that both *buru* and *jom* have an equivalent distribution: each can combine with the verbal affixes and enclitics, as in the series *-ke-d-a* ‘completive, transitive, indicative’: each can be the root of the predicate, in clause-final position, but each can also fill the (object) argument slot before the verb, and host the subject clitic *=ko* ‘they two’.

16. For a broadly similar discussion of these points see Croft (1991, 2000, 2001), particularly the latter. More detailed similarities and differences in our treatment will be mentioned in more detail below.

17. Cf. the critique in Croft (2001: 30–32) of “cross-linguistic methodological opportunism” – of just using that small subset of tests that fit a particular point to be proven. At the same time, we must confess that the ideal – to test all possible distributional features of the candidate classes – is too large an undertaking to be practical in this article, and we confine ourselves to a canonical subset of distributional tests.

We use the term "prima facie distributional equivalence" to flag a common short-cut in the application of this principle. Rigorously applied, all members of both putative classes should be equally acceptable in both primary syntactic functions as argument and predicate. However, in the flexible word class literature (including on Mundari) one often finds a greater burden of exemplification falling on the use of all types of words as predicate, with much less attention to the converse situation where a range of words are tested for acceptability in argument position.[18] As a result, far more discussion of semantic compositionality concerns predicate uses, so that we hold off on the discussion of argumental uses until Section 3.3, where we return to the rubric "bidirectionality".

3.2. *Compositionality*

The criterion of compositionality – ANY SEMANTIC DIFFERENCES BETWEEN THE USES OF A PUTATIVE "FLUID" LEXEME IN TWO SYNTACTIC POSITIONS (SAY ARGUMENT AND PREDICATE) MUST BE ATTRIBUTABLE TO THE FUNCTION OF THAT POSITION[19] – ensures that the meaning of the composed word must be pre-

18. For another critique on overreliance on predicative contexts in arguing against word class distinctions, see Croft (2001: 84).

19. We may apply this principle equally to affixal material, rephrasing it as follows: COMBINATION OF MEMBERS OF TWO POSSIBLY DISTINCT CLASSES WITH A DERIVATIONAL OR INFLECTIONAL ELEMENT HAVING THE SAME FORM NEED NOT ARGUE AGAINST WORD-CLASS DIFFERENTIATION, AS LONG AS THE DIFFERENCES IN DENOTATIONAL EFFECT PRODUCED BY THAT ELEMENT CANNOT BE ATTRIBUTED TO INTERACTIONS OF ITS SEMANTICS WITH THAT OF THE BASE. This reflects the fact that many languages re-use formally identical material with more than one class, though with differing semantic effects depending on the class they combine with – cf. the Indonesian prefix *ber-*, which can combine with noun or verb bases with quite different effects (basically 'have N' with nouns but no clear semantics with some verbs and reflexive semantics with others). The double use of person-number affixes in many languages, marking possession when attached to nouns but encoding a core argument (typically subject) when attached to verbs, is another widespread example.
This principle is applicable to one important affix in Mundari: the infix ⟨*pV*⟩ (where V copies the preceding vowel). This infix is found, in identical form, with words that would be rendered in most languages by two word classes: transitive verbs and adjectives. Pairs indicating the former possibility, where it indicates reciprocal action (Osada forthcoming), include *erang* 'scold', *e⟨pe⟩rang* 'quarrel'; *lel* 'see', *le⟨pe⟩l* 'see each other'; and *ad* 'miss', *a⟨pa⟩d* 'miss each other', while pairs illustrating the second include *marang* 'large', *ma⟨pa⟩rang* 'very large'; *huRing* 'small', *hu⟨pu⟩Ring* 'very small'; and *jiling* 'long', *ji⟨pi⟩ling* 'very long'. Now the combining of a single infix across all these words could be chalked up as a point in favour of all forms belonging to a single word class here. However, it is difficult to derive reciprocal semantics explicitly and completely from intensification (notwithstanding the occasional parallel in other languages where there is some formal similarity between reciprocal and intensive forms, e.g., Arabic and Tigrinya), or vice versa, so this fact cannot be taken as support for a monocategorialist position. In fact, Anderson & Zide (2001) suggest that these two infixes have different diachronic origins within Austroasiatic.

dictable from the meaning of its parts plus the meaning contributed by syntactic function or constructional frame (e.g., predicate).[20]

For example, in the celebrated Nootka examples in (21),[21] originally from Swadesh (1939), their Straits Salish equivalents in (22) (Jelinek 1995), and their Tagalog equivalents in (23) (cited in Sasse 1993), the difference between the predicate use, meaning 'dances, is dancing', and the argument use, meaning 'dancer; the one who is dancing' is attributable to the semantic functions of predicating and referring, respectively, that are linked to the predicate and argument (or VP, and NP) positions.

(21) a. *mamuːk=ma* *quːʔas-ʔi*
working=PRES.INDIC man-DEF
'The man is working.'

b. *quːʔas=ma* *mamuːk-ʔi*
man=PRES.INDIC working-DEF
'The working one is a man.'

(22) a. *čəy=Ø* *cə* *s-wəy'qə'*
work=3SG DET STAT-male
'He works, the (one who is a) man.' (= The man works.)

b. *s-wəy'qə'=Ø* *cə* *čəy*
STAT-male=3SG DET work
'He is a man, the (one who) works.' (= The working one is a man.)

(23) a. *nagtatrabaho* *ang* *lalaki*
work.AT.IMPF TOP man
'The man is working.'

b. *lalaki* *ang* *nagtatrabaho*
man TOP work.AT.IMPF
'The one who is working is a man.'

All of these examples satisfy the compositionality criterion, which is a key reason why Nootka, Salish, and Tagalog have become the classic examples of languages challenging the noun-verb distinction.

---

20. Our formulation is less demanding than that in Croft (2001: 67) and the formulation suggested by an anonymous *LT* referee, both of whom would also want to exclude the semantic effects of "coercion" to a particular syntactic function, particularly to participant in the nominal use (dancer < dance). The issue here is whether we want to get the semantics from the construction, in particular the predicate or argument position, or from the lexeme itself. We prefer to leave open a role to syntactic structure in contributing to meaning, though subject to the constraints we spell out below. And since our (weaker) requirement is already sufficient to deal with the Mundari case, as shown in the rest of the paper – so that the stronger requirement would a fortiori deliver the same results – we believe our formulation is sufficient.
21. We return below to more subtle differences between 'work' and 'man' in Nootka.

And there are some argument–predicate pairings of Mundari lexemes that at first sight seem to rather comparable to these examples, if one is willing to admit the contribution of aspect markers suffixed to the predicate in composing its meaning.

Consider the following three pairs: in (24a) *dasi* heads a referring expression with the meaning 'servants', while in (24b) it occupies the predicate slot and bears the "initiated continuous" aspectual[22] marker *-aka-*, with the meaning 'are (since earlier) working as servants'. Likewise in (25a) *baRae* 'blacksmith; member of blacksmith caste' heads a referring expression, with the meaning 'blacksmith', while in (25b) it occupies the predicate slot: it is not outrageous to argue that the meaning 'become a blacksmith, enter the blacksmith caste' derives from the interaction of the referring meaning of *baRae* with the initiated continuous aspect (this caste is considered by the Munda to be beneath them, and the sentence would be used of a man who moved down a caste through marriage or sexual union with a *baRae*-caste woman; once such hypogamy occurs, the downgrading in caste is irreversible). In (26a) the loanword *mastaR*, from English via Hindi, is used in a referring expression, while in (26b) it is used as a predicate and, together with the "initiated continuous" aspectual suffix, has the meaning 'work as a teacher'.

(24) a. *dasi-ko=ko* *kami-ta-n-a.*
servant-PL=3pl.S work-PROGR.OR-INTR-INDIC
'The servants are working.'

b. *dasi-aka-n-a=ko*
serve-INIT.PROG-INDIC=3pl.S
'(They) are working as servants.'

(25) a. *baRae-ko=ko* *susun-ta-n-a*
blacksmith-PL=3pl.S dance-PROGR.OR-INTR-INDIC
'The blacksmith caste members are dancing.'

---

22. Mundari has a complex aspectual system, with four distinct forms in the perfective series, and two in the imperfective series. Here we use somewhat different glosses than those to be found in Munda (1971) and Osada (1992: 94–97). For the four members of the perfective series, the options are *a* (Munda's 'cislocative', referring to completed actions remote in time, *ke* 'completive', which simply marks completion of an action without reference to any other action, *le* 'anterior', which marks an action completed before some other action, and *ja* 'inceptive', which marks the inception of an action. For the two members of the imperfective series, the options are *ta* (PROGR.OR for 'progressive oriented') which marks an action in progress but oriented to some future endpoint and *aka* (initiated progressive) which marks a situation now in force, but focussing on the fact that this current state of affairs has already been initiated. Contrasting the last two, compare *dubtanako* 'they are in the process of sitting down', which is ongoing but oriented to the endpoint of reaching a sitting state, and *dubakanako* 'they are sitting', where the state of sitting has already been initiated.

b. *soma=eq* *baRae-aka-n-a*
Soma=3sg.S *baRae*-INIT.PROG-INTR-INDIC
'Soma has become a *baRae* [lower caste member].'

(26) a. *mastaR* *isTuDeNT-ko=eq* *paRao-ke-d-ko-a*
teacher student-PL=3sg.S teach-COMPL-TR-3pl.O-INDIC
'The teacher taught the students.'

b. *soma=eq* *mastaR-aka-n-a*
Soma=3sg.S teacher-INIT.PROG-INTR-INDIC
'Soma is a teacher, is working as a teacher.'

With appropriate analysis of the intricate aspectual system, and an account of the contribution of dynamic aspects to the derivation of meanings like 'work as a servant', 'become a *baRae* caste member', and 'work as a teacher', examples like these seem to satisfy the compositionality criterion.

But we now consider two types of difficulty, whose resolution requires us to introduce a corollary of the compositionality requirement, that of "compositional consistency": THERE SHOULD BE ISOMORPHIC SEMANTIC CHANGES IN ALL LEXEMES PLACED IN A GIVEN FUNCTIONAL POSITION.

Monocategorialists wanting to analyse examples like (24)–(26) above typically appeal to an argument of "coercion" from the constructional slot – i.e., the "extra" semantics is argued to fall out from the function of the syntactic slot (see, e.g., Langacker 1987). But if the extra semantics is indeed attributable to the constructional slot, then all semantically comparable words in the same slot should undergo the same semantic augmentation. For example, if the nominal uses of 'eats', 'drinks', 'smokes' (i.e., things eaten, things drunk, things smoked) are to be derived from the verbal meanings by this argument, then the evidence can only be used to argue for a single word class if it applies to all other comparable words – but one can't say, in English, e.g., 'inhales' or 'sniffs' for 'things inhaled' or 'things sniffed'.

A complication to applying this corollary comes from the fact that aspectual information on the predicate slot may interact with the predicate lexeme to produce some differences in semantic contribution, or in the acceptability of the new predicate with different aspects. For example, when *mastaR* occupies the predicate slot it requires the "initiated progressive" aspect suffix *-aka*, whereas when *baa* 'flower' occupies the predicate slot it combines with the "progressive oriented" suffix *-ta*, as in (27).

(27) *ne* *daru=eq* *baa-ta-n-a*
DEM tree=3sg.S flower-PROGR.OR-INTR-INDIC
'That tree is flowering.'

A monocategorialist could then attempt to attribute this difference to subtle interactions between these aspect types and the nature of the events being de-

picted: working as a teacher or becoming a *baRae* involve an initial transition (entry into the profession or caste followed by a steady activity), while flowering involves a build-up of steps (budding, first buds bursting into flower, gradual extension to all buds on the tree) that, though continuous, are cumulative steps oriented towards an endpoint of the whole tree being in flower. To evaluate this counter-argument we would need a far better understanding of the complexities of Mundari aspect than we currently have, so we will err on the side of generosity in attributing such subtle differences in semantic increments to interactions of aspect with the Aktionsart of the depicted process. An amended version of our corollary which would tolerate some minor variations in increment, therefore, is THAT THERE SHOULD BE THE SAME SEMANTIC CHANGE IN ALL LEXEMES PLACED IN THAT POSITION, EXCEPT FOR SEMANTIC INTERACTIONS ATTRIBUTABLE TO INFLECTIONS BORNE BY IT, E.G., ASPECT. This more liberal attitude, however, does not let the monocategorialist completely off the hook when we push the Mundari data a bit further.

First, one would still need to find an aspect allowing *mastaR*, *baRae*, *baa*, etc. to be used in the exactly composed meaning 'be a teacher', 'be a blacksmith', 'be a servant', etc. But to express these concepts, which are the ones most comparable to the Nootka, Salish, and Tagalog examples above, a different construction with the copula *tan* is used, rather than employing the lexical item directly in the predicate slot.[23] For example, the word *hoRo* 'person;

23. Mundari has two copula constructions: one, with *tan*, for identity (equative or ascriptive), and another, with *menaq*, for locative or existential clauses, along with its corresponding negatives *bang* (3NON-SG.NEG), *banoq* (3SG.INAN.NEG) and *banggaiq* (3SG.AN.NEG), though the contrast is confined to the present tense (see also Munda 1971). Both are confined to nominal complements: only nouns and locative expressions, not verbs, can be their complements. The following examples illustrate their contrasting uses:

(i) *Soma tan-iq*
Soma be=3sg
'It is Soma.'

vs.

(ii) *Soma oRaq-re menaq-i-a*
Soma house-LOC be.located-3sg-INDIC
'Soma is in the house.'

We note one special modern development, based on a Hindi calque, that constitutes an exception to the unavailability of copulas with verbs: expressions of obligation of the type "X has to V" can be formed by putting the subject in the dative and adding *-menaq* directly to the verb, as in

(iii) añ-ke senoqmenaq
1SG-DAT go-BE.LOCATED
'I have to go.'

Munda person' can only be used to predicate the meaning 'be a Munda' if it is combined with a copula, as in (28a); if it is placed directly in the predicate position it adds the semantic increment 'speak [Munda] language', as in (28b).

(28) a. *ne dasi hoRo tan-iq*
this servant Munda COP-3sg.S[24]
'This servant is a Munda.'
b. *ne dasi hoRo-a=eq*
this servant speak.Munda-INDIC=3sg
'This servant speaks Munda.'

Similarly, consider how one says 'these are servants'. A special construction, taking the form Subj Compl Copula(=Subj.Clitic) must be used, as in (29a).[25] But now the copular construction poses problems for the equivalent combinatorics criterion (Section 3.1 above), since it is not available with words which would be prototypical verbs in other languages, like *hijuq* 'come': (29b), for example, is ungrammatical.

(29) a. *en hoDo-ko munDa=ko tan=ko*[26]
those man-PL headman=3pl.S be=3pl.S
'Those men are headmen.' (Langendoen 1967: 84)
b. **niku / hijuq-tan=ko*
these come-COP=3sg
'These are coming.'
c. **Soma hijuq-tan-iq*[27]
Soma come-COP-3sg
'Soma is coming.'

The second difficulty comes from the fact that the above examples – which are as close as we get to showcases for a monocategorial analysis – are by no means typical of the whole lexicon. Looking across a wider range of lexemes,

---

parallel with Hindi *mujhe jaanaa hãī*). The dative subject, uncharacteristic for Munda, also bears witness to the calqued nature of this construction; see Osada (1999).

24. The 3rd singular subject allomorph *iq*, instead of *eq*, is used after the copula and after relative clauses.
25. For further details on the Mundari copula see Langendoen (1967) and Osada (1992), though Langendoen's analysis differs from ours on some points, e.g., by counting the verb *tai* as a copula (see, e.g., his example 22), whereas we regard it as a regular intransitive verb, meaning 'remain'.
26. For some unclear reason, not all younger speakers accept this construction with plural subjects; speakers of all ages accept this construction with singular subjects. Note that this example is cited in a (transliterated) form of the original example in Langendoen (1967), who writes 'man' with a retroflex stop instead of a flap.
27. Note, though, that the homophonous string *hijuq-ta-n-iq* [come-PROGR.OR-INTR-3sg] is acceptable as a relative clause meaning 'who has come'.

it is common for the semantic difference between argumental and predicate uses to way exceed that attributable to the syntactic position, or the small perturbations due to interactions with the aspectual system.[28] In such cases we are either dealing with lexical derivation by zero conversion, or with a Broschartian language with specific types of semantic agreement according to the semantic class of the root.

Let us give some examples from transitive uses. A common meaning for basically nominal roots used as transitive verbs is 'cause a(n) N to exist': examples are *bir* 'forest; plant a forest', *lad* 'pancake; make pancakes', *maNDi* 'food; make food'. But frequently conversions of this type take on an additional metaphorical meaning that can no longer be precisely paraphrased as causatives of existence. In the case of (1) and (2), for example, repeated here as (30a, b), the semantic increment 'gather (so as to resemble a …)' does not mean simply 'cause to be a mountain', or at most 'cause to become a mountain', but means more specifically 'to heap up'. Even though the metaphor it appeals to is rather obviously based on a caused existence meaning, by likening a large group to a mountain, it is nonetheless one specific semantic addition, instead of other imaginable additions (e.g., 'cause to be tall', 'cause to be outstanding'), and must therefore be treated as lexicalized. As further support for the arbitrariness of this increment, note the different effects on the equivalent noun in Sora, namely *baru:* 'a hill, forest', whose corresponding transitive verb is *baru:* 'to make a clearing on the slope of a hill in order to grow dry crops thereon'.

(30) a. *buru=ko* *bai-ke-d-a*
mountain=3pl.S make-COMPL-TR-INDIC
'They made the mountain.'

b. *saan=ko* *buru-ke-d-a*
firewood=3pl.S "mountain"-COMPL-TR-INDIC
'They heaped up the firewood.'

If we look at other transitive-predicate uses of words that are, crosslinguistically, typically nouns, we find again rather a wide range of semantic increments. Taking our cue from (30b), we might expect the basic pattern to be 'cause a(n) N to exist'. Examples are *bir* 'forest; plant a forest', *lad* 'pancake; make pancakes', and *maNDi* 'food; make food'. But we also find various pairs where the semantic addition is 'acquire N', i.e., 'cause N to be in one's possession'; an example is *sim* 'fowl; acquire fowls'.[29] And with a few words, the

28. Cf. Neukom (2001: 16) on Santali: "The relationship between the meaning as argument and the meaning as predicate is not obvious, e.g. *bəhu* means 'bride', in argument position, but 'take a bride for somebody' in predicate position".

addition seems to be 'do (to OBJ) as an N does': thus *gaRa* 'river; dig (like a river)'.

The analyst might be tempted to respond to the range of semantic increments when nouns are used in transitive constructions – 'cause (a)n N to exist', 'acquire N', 'do (to OBJ) as a N does' – by proposing a Broschartian analysis, in which there is a small but predictable set of increments that is determined by the ontological class of the object. If that were the case, we should be able to hold the lexical subclass constant, and always get the same increment when using it in a transitive frame. But now consider what happens if we examine the use of a range of 'tool' lexemes in transitive constructions. For some, we get the 'instrumental' increment 'act upon (OBJ) using an N': thus *laTab* 'scissors; cut with scissors' (31). For others, we get the 'manufacturing' increment 'make into a N': thus *kaTu* 'knife; forge into a knife' and *aq-sal* 'bow and arrow; make a bow of something, to turn something into a bow, to call something bow'. Revealingly, in each of these last two cases there is a formally distinct verb denoting the instrumental activity: *had* 'cut with a knife', *ToTeq* 'shoot with an arrow', suggesting that transitive uses will receive an instrumental reading if no special instrumental lexeme is available, but will otherwise receive a manufacture meaning. But there are some lexemes that allow both the manufacture and a specialized instrumental meaning, with a distinct form for the general instrumental meaning: *kapi* 'axe; form into a hunting axe-head, to strike somebody so or so many times with one's hunting axe' alongside *maq* 'cut with an axe'.

(31) *soma* *kaTa-re=q* *laTab-ja-n-a*
Soma leg-LOC=3sg.S cut.with.scissors-INCEP-INTR-INDIC
'Soma cut his leg with scissors.'

To sum up this section, when what we would expect to be nouns are used, both in intransitive and transitive constructions, the semantic increment is far from constant. Even though some small perturbations may be attributable to interactions with aspect, many problems remain. It is difficult to account for larger perturbations ('work as N' and 'speak [language] N' with intransitives, and 'cause to be(come) N', 'acquire N', and 'use N' with transitives). And one needs to account for the fact that an alternative copular construction is needed to use nouns simply as predicates without adding any further aspectual information. Even if the analyst tries to adopt a Broschartian analysis by breaking

29. Hoffmann's sentence *simkedkoale* 'we have acquired ("fowled") them (fowls)' is rejected by our informant Maki Purti as ungrammatical, but she accepts the form *simkedale* without the object marker *ko* (though with the transitive marker *-d-*). The interaction of cognate object verbs like this with object agreement needs further investigation.

the lexicon up into detailed classes, in the hope that at least at that level the semantic increment stays constant, they are still confronted with major variations in semantic increment even within a single lexical class, like 'instruments'.

Overall, then, it is clear that the semantic increments that appear when entity-denoting terms are used in the predicate slot are significantly irregular and non-compositional – something one would expect in a zero conversion language, but not in the other types. The fact that some nouns, when they become verbs, have only minor semantic additions, perhaps accountable for by a sufficiently ingenious analysis of the aspectual system, needs to be weighed against the fact that many other nouns undergo much more serious additions. To argue from just a few favoured cases that there are no word class distinctions is like arguing that English lacks word class distinctions because there are a few fluid lexical items – of the type 'kiss', 'whore', and 'flower' – while disregarding either the much greater semantic differences between nominal and verbal use, or the complete unavailability of conversion, with the vast majority of other items.

3.3. *Bidirectionality*

So far we have been concentrating on the use, as predicates, of words which can also function as arguments – or, more precisely, as the head of phrases (NPs) which function as arguments. However, to establish that there is just a single word class, it is not enough for Xs to be usable as Ys without modification: it must also be the case that Ys are usable as Xs. In the history of arguments about single-word class languages, a decisive counter-attack against Swadesh's (1939) monocategorialist position came when Jacobsen (1979), and subsequently Schachter (1985) and Anderson (1985), pointed out that even though either 'work' or 'man' could fill the predicate slot in Nootka, and even though both 'work' and 'man' could fill the (subject) argument slot once combined with a determiner (see (21a) above), only 'man' can fill this slot without a determiner (Y). This rather subtle evidence against bidirectional equivalence was then taken as evidence for a (rather weak) distinction between nouns and verbs in Nootka: nouns are words that can be used as arguments without determiners, whereas both nouns and verbs can be used as arguments with determiners, and both nouns and verbs can be used directly as predicates.

(32) a. *mamuːk=ma* *quːʔas*
working=textscpres.INDIC man
'A man is working.'

b. **quːʔas=ma* *mamuːk*
man=PRES.INDIC working
'A working one is a man.'

In the Mundari case, it is common for primary nouns to be used as predicates, though as we have already seen there are many semantic complications. We have also established, so far with just the single example of the prima facie verb *jom* 'eat' used as a noun ('food'), that it is possible to use at least some verbs directly in the argument slot; this example is repeated here as (33a, b, c). A further example is (33d, e), which first illustrates the predicate use of *dal* 'beat, hit' (33d), then its argumental use to mean 'a beating with a stick' (33e).

(33) a. *maNDi=ko jom-ke-d-a*
food=3pl.S eat-COMPL-TR-INDIC
'They ate the food.'

b. *jom=ko nam-ke-d-a*
food=3pl.S get-COMPL-TR-INDIC
'They got the food.'

c. *jom=eq nam-ke-d-a*
food=3sg.S get-COMPL-TR-INDIC
'(S)he got the food.'

d. *hon-ko=eq dal-ke-d-ko-a*
child-PL=3sg.S beat-COMPL-TR-3pl.O-INDIC
'(S)he beat the children.'

e. *mid DaNDa dal=le nam-ke-d-a*
one stick beating=1pl.EXCL.S get-COMPL-TR-INDIC
'We got one stroke of beating.'

How common, and how syntactically thoroughgoing, are such cases? (Recall that Mundari nouns, when used as arguments, take no affixes, except for role-marking postpositions in oblique functions, that they are able to constitute a complete NP without any determiner, or may be preceded by a determiner, and that the last NP before the verb typically hosts the subject pronoun as a clitic.) Do they represent genuine monocategoriality, comparable to the Salish or Tagalog cases, or are they simply sporadic cases of zero conversion comparable to the the nominal use of 'drink' in English 'Did you remember the drinks'?

In fact, though we can use some verbs freely as arguments, the vast majority must effectively be converted into headless clauses before being placed in an argument slot. In the case of *jom* in (33b) and *dal* in (33e), the lexeme is placed directly into a slot appropriate for a noun, either as the sole word in the NP in (33b), or as the head of the NP in (33e), modified appropriately. Most verbs, by contrast, and all adjectives, can only be placed into an argument slot if they are followed by appropriate aspectual and transitivity markers, and where the referent is 3rd person singular, they must be followed by a special form of the agreement affix, namely *-iq* instead of *-eq*, which is effectively a subordinator. (34)–(35) illustrate this construction for the verbs *om* 'give' (34) and *susun* 'dance' (35).

(34) a. *om-ke-n=iq* *goeq-ja-n-a*
give-COMPL-INTR=3sg.S die-INCEP-INTR-INDIC
'the one who gave died'
b. **om=eq* *goeq-ja-n-a*
give=3sg.S die-INCEP-INTR-INDIC
c. **om=iq* *goeq-ja-n-a*
give=3sg.S die-INCEP-INTR-INDIC
d. *om-ke-d=iq* *goeq-ja-n-a*
give-COMPL-TR=3sg.S die-INCEP-INTR-INDIC
'the one who was given to died'

(35) a. *susun-ta-n=iq* *landa-ja-n-a*
dance-PROGR.OR-INTR=3sg.S laugh-INCEP-INTR-INDIC
'The one who is dancing has laughed.'
b. **susun=iq* *landa-ja-n-a*
dance=3sg.S laugh-INCEP-INTR-INDIC
c. **susun=eq* *landa-ja-n-a*
dance=3sg.S laugh-INCEP-INTR-INDIC

Similar behaviour for adjectives, which display the same restrictions as for verbs in this regard, is illustrated in (36) with the adjective *marang* 'big'.

(36) a. *marang-ke-n=iq* *goeq-ja-n-a*
big-COMPL-INTR=3sg.REL die-INCEP-INTR-INDIC
'the one who was big died'
b. **marang=iq* *goeq-ja-n-a*
big-3sg.REL die-INCEP-INTR-INDIC
c. **marang=eq* *goeq-ja-n-a*
big=3sg.S die-INCEP-INTR-INDIC

The restrictions just noted apply to verbs used in prototypical argument position, which are the most rigorous testing ground for claims of bidirectional distributional equivalence. When used in complement clauses, by contrast, they do not need to take aspectual or transitivity markers, and can appear directly as the sole element of an NP, as in (37) and (38).

(37) *her=ko* *caba-ja-n-a*
sow=3pl finish-INCEP-INTR-INDIC
'They have finished sowing.'

(38) *dub=ko* *laga-ja-n-a*
sit=3pl be.tired.of-INCEP-INTR-INDIC
'They are tired of sitting.'

In such infinitive-like complement positions, then, we do not find the same idiosyncratic restrictions on which lexemes can occur that we found in the case of *jom*. However, this is hardly a distinctive characteristic of monocategorial languages, since it is well established that many languages blur the distinction between nominal and verbal characteristics in such non-prototypical contexts (cf. Hopper & Thompson 1984): infinitives often exhibit mixed nominal and verbal characteristics, just as generic or incorporated nouns often exhibit a restricted range of nominal characteristics. It is thus important to apply bidirectional tests to prototypical functions, such as the use as subject arguments in (34) and (35), rather than peripheral clausal functions like purpose or other complements.

We can thus distinguish verbs with deverbal conversions, like *jom* 'eat; food', which can be used as arguments with equivalent combinatorics to nouns, from verbs which require special morphosyntactic treatment before they can function as arguments; *om* is an example.[30] So although all verbs can indeed be used directly in "clausal argument" positions, as in (37) and (38), it is the unavailability of all but a limited subset of verbs for direct prototypical argument use, as in (34) and (35), that shows the absence of real bidirectionality in Mundari.

3.4. *Exhaustiveness*

The principle of exhaustiveness states that it is not sufficient to find a few choice examples which suggest word class flexibility. Since word classes are partitionings of the entire lexicon, equivalent statements need to hold for all relevant words in the lexicon that are claimed to have the same class. In recent publications, Croft and Baker (among others) have made similar points:

> How do we know that when we read a grammar of an obscure "flexible" language X that the author of the grammar has systematically surveyed the vocabulary in order to identify what proportion is flexible? If English were spoken by a small tribe in the Kordofan hills, and all we had was a 150 page grammar written fifty years ago, might it look like a highly flexible language? (Croft 2001: 70)

> An important typological difference exists only if categorial ambiguity extends to an entire open class of inflectionally similar words, thereby affecting the overall grammar of the language. (Baker 2003: 177)

In practice, since it is difficult to check every one of tens of thousands of lexical items, we at least need a large enough sample[31] that it would plausibly pick

30. There are further restricted contexts in which the verb can be used as a syntactic argument, e.g., *en dub-ke=ñ dub-kena* 'I sat and sat for a long time (in the plane)', lit. 'that sit-ASP I sat', though even here the aspectual enclitic *-ke* is present.

31. An anonymous *LT* referee, echoing Croft's (2001: 70) question "How many are enough?",

up limits in productivity, and that includes several representatives of as many denotational classes as possible.

Many of the influential discussions of word classes in particular languages do not meet this criterion. For example, Jelinek's (1995) analysis of Salish only considers a handful of lexical items. This is not to say they are wrong, merely that they have yet to supply conclusive evidence, and are at risk of having presented selective data. We will argue that the claimed fluidity of Mundari looks less productive once the full range of data is considered, and in fact Hoffmann himself retreated from the original position (stated in his 1903 grammar) once he was forced, in his gigantic masterpiece, the *Encyclopaedia Mundarica*, to make an exhaustive analysis of thousands of lexemes.

More extensive checks of this type throw up problems for each of our criteria above: distributional equivalence, with its corollary of bidirectionality, and compositionality. We have already mentioned some problems of this type in Sections 3.2 and 3.3. In this section we push our investigations further by first looking at a semantically targeted sample of lexical items chosen to represent four lexical domains – proper nouns, animal names, plant names, and kinship terms – and then giving statistics on the combinatorial possibilities of two samples from the lexicon – a smaller one included as an Appendix to this article, and a larger one of around five thousand lexical items carried out by the second author (comprising about 10 % of the attested Mundari lexicon) for which the results are available on the web at http://munda.chikyu.ac.jp/EM/.

3.4.1. *Nominal subclasses of limited availability for predicate use.* First, PROPER NAMES, such as *Ranci* 'Ranchi', are unavailable for predicate use (39a), and cannot even be combined directly with the copula (39b). Instead, one needs to place a postposition after the proper noun, then add the copula (39c).

asks whether we would demand 100% conformity. We would, subject to three caveats: (i) obviously there is a difference between what we would demand for an ideal proof, and what can be taken as demonstrated of a given language at a given history in its investigation, since checking all distributional contexts for every lexeme is an immense undertaking; (ii) since our statements are made about the major open classes, we have no problem – in contradistinction to Croft (2001: 70) – with removing small numbers of lexemes into minor word classes (which by definition are finite) where they display distinct combinatorics and/or semantic effects; (iii) we also assume that the major word classes will be further divisible into subclasses (e.g., proper nouns, verbs grouped by aspectual class and so on) and that these will be mirrored in distributional differences at a more specific level.

(39) a. **añ-aq oRaq ranci-ta-n-a*
1SG-GEN home Ranchi-PROGR.OR-INTR-INDIC
'My home town is Ranchi.'
b. **añ-aq oRaq ranci menaq*
1SG-GEN home Ranchi LOC.COP
'My home town is Ranchi.'
c. *añ-aq oRaq ranci-re menaq*
1SG-GEN home Ranchi-LOC LOC.COP
'My home town is Ranchi.'

Second, it is common in many languages for just a small subset of KINSHIP TERMS – typically just 'mother' and 'father' – to permit verbal as well as nominal uses, though even there the semantic increment may be different, with the 'mother' word meaning 'nurture, care for as a mother does' and the 'father' word meaning 'beget' (cf. Evans 2000b). Even these two terms, then, violate the compositionality requirement, but in addition a consideration of the full kin term set often reveals that conversion is non-productive: 'uncle', 'sister', and 'cousin', for example, are not available as verbs in English, outside the very special 'reported vocative' use: 'Don't you "uncle" me!" (i.e., don't call me uncle).

It is clear that any claim, for a language like English, that kinship terms were precategorial because 'mother' and 'father' can occur in both nominal and verbal slots, would need to be backed up by demonstration that other kin terms exhibited similar fluidity, and here it would quickly founder against a more exhaustive sampling of the data.

With some minor differences of detail, the Mundari facts are remarkably like English. A small set of nouns in the domain of kinship have verbal uses, and the semantic increment is irregular. Thus the nouns *engga* 'mother', *haga* 'brother', and *geRe* 'man's sister's child' can be used as verbs; in the case of 'mother' the lexicalized semantic addition is minimal, but in the case of the other two it is substantial. With *haga* 'brother', for example, it is based on an extension of brotherhood to all members of the same clan (and note that (41) could still be used between women; the root *misi* 'sister' could not be substituted).

(40) *engga-oq-ta-n-a=eq*
mother-PASS-PROGR.OR-INTR-INDIC=3sg.S
'She is becoming a mother.'

(41) *ale do in-ku-loq=le*
we.EXCL TOP that-PL-with=1pl.EXCL
haga-ta-n-a
brother-PROGR.OR-INTR-INDIC
'We are in the same clan as them.'

(42) *keoRa geRe-oq-ta-n-a*
Keora maternal.village-PASS-PROGR.OR-INTR-INDIC
'Keora has become my maternal village (following the remarriage of my mother to someone of that village).'

But most kinship terms are nouns only: none of *apu* 'father', *misi* 'sister', *boko* 'younger sibling', *baDa* 'elder uncle', *baDi* 'elder aunt', *kaka* 'younger uncle', and *kaki* 'younger aunt', for instance, can be used as verbs:

(43) **apu-oq-ta-n-a=eq*
father-PASS-PROGR.OR-INTR-INDIC=3sg.S
'He is becoming a father.'

Third, most NAMES OF ANIMALS cannot be used as predicates. Examples include *seta* 'dog', *pusi* 'cat', *uRiq* 'cow', *tuRu* 'squirrel', *sukuRi* 'pig', *merom* 'goat', *miNDi* 'sheep', and *sim* 'hen', though the behaviour of this last term is more complex – Hoffmann gives examples of it used as a transitive verb, with the meaning 'acquire fowl' (see above) while Maki Purti rejected this use.

One of the few animal terms permitting predicate use is *kula*, which can be used, though only when followed by the passive, with the meaning 'turn into a tiger' (in the way people become vampires in European mythology):

(44) *soma=eq kula-oq-ta-n-a*
Soma=3sg.S tiger-PASS-INCEP-INTR-INDIC
'Soma has become a tiger.'

For the realm of animal nouns, then, Mundari is less productive in its possibilities for predicate use than English, which has, for example, *to dog*, *to fish*, *to bitch*, *to chicken (out)*, *to snake*, and *to pig (out)*, though not **to giraffe*, **to lizard*, **to dingo*, or **to deer*.

Fourth, most PLANT TERMS cannot be used as predicates, including *kaNTaRa* 'jackfruit' (**kaNTaRa-tana*), *uli* 'mango' (**uli-tana*), and *kadal* 'banana' (**kadaltana*). One of the few plant names that may be used as a verb is *jojo* 'tamarind', which can be used as a verb with the meaning 'to be sour'.

This consideration of four semantic subclasses shows quite clearly that, though for most nominal subclasses (except proper nouns) it is possible to find a couple of lexical items that can be used as predicates, this is a property of specific lexical items and in no way generalizes across all members of the categories, as it should in an omnipredicative, precategorial, or Broschartian language. In fact, the impossibility of certain nouns being used as verbs was noted by Hoffmann himself: the entries in his *Encyclopaedia Mundarica* for *ade* 'ginger plant', *ambuRu* 'Indian hog-plum', *angkusi* 'hook, plough', *awa* 'kiln for baking roof tiles', and *apu* 'father', for example, state explicitly that these can only be used as nouns. It is significant that this more restrictive view

of Mundari fluidity is found in the *Encyclopaedia* which is both later and more exhaustive than his grammar: it represents both a more considered view, and one that must account for the behaviour of every lexeme, rather than a select few.

3.4.2. *Figures from small sample.* To give a more quantitative picture of the possibilities of NOUN-TO-VERB and VERB-TO-NOUN conversion, we now give figures compiled from a sample of 105 lexical items, listed in the Appendix. Our sample was designed to include a range of ontological types, including all those covered by Broschart for Tongan: in addition to a range of terms for states and processes, and positions, it included terms for stages (e.g., old man, boy), sex (e.g., man/male, woman), material (e.g., stone, dust), qualities (e.g., bad, good), colour (e.g., red), kinship relation (e.g., father, mother), social domain (family, clan, tribe), body part (e.g., head, foot), instrument (e.g., hammer, axe), product (e.g., pancake, song), manner (very, slow, fast), value/quantity (one, two, part), task (cowherd, smith), time of day (morning, evening), place and geographic features (inside, mountain, river), natural kinds including both plants and animals, provisions (food, drink), days and festivals, names, nationalities and tribes. It also included the key cases discussed by Hoffmann and Bhat, and the words taken from the first two sentences of a sample text about harvesting.

For the purposes of the figures below, lexical items allowing more than one function were classified as basically nominal or basically verbal following the analytic decisions in Hoffmann regarding order of appearance in the lemmas of his dictionary, supplemented by considerations of semantic inclusion, such that if one meaning includes the other in its definition (e.g., 'acquire fowl' includes 'fowl') then it is counted as more basic. Of course the right direction of derivation could sometimes be disputed, but this would only affect the figures regarding direction of conversion, not those regarding whether conversion can occur. Nonetheless, in a few cases where the question of directionality is particularly difficult (*rimbil* 'cloud/cloudy', *durang* 'sing/song', *soan* 'stink; smell') these were not included in the figures. Alongside each term we indicate whether the closest English equivalent allows a parallel conversion.

Overall, of the 105 lexical items that include at least a nominal or a verbal use, 74 are convertible, i.e., a little under three quarters. For the 41 basic verbs in our sample, 27 allow nominal use (i.e., around 65%) and 14 do not not; the corresponding figures for the English translation equivalents that were verbs were slightly higher: 29 allow nominal use, 14 do not. For the 64 basic nouns in our sample, 47 can function as predicates, i.e., 74 %. This is a little higher than the comparable English figure of 65 %. Overall, then, around 72 % of the lexical items can function either as nouns or as verbs, a figure slightly higher than the corresponding figure for English, but well short of the 100 %

Table 1. *Figures for noun-only, verb-only, and noun-verb for a sample of 3,824 lexemes*

	Number	Percentage
Noun only	772	20
Verb only	1099	28
Both noun and verb	1953	52
Total	3,824	100

that would be required to establish the lack of word class distinctions in the language.

3.4.3. *Figures from large sample.* As a double check on the figures in Section 3.4.2, and also because it was biased towards finding word class flexibility by the inclusion of all sample lexemes discussed by Hoffmann and Bhat, the second author examined a much larger sample of more than 5,000 entries, comprising around 10% of the items in *Encyclopaedia Mundarica*. These were selected by entering all the entries from pages chosen at random throughout the dictionary; out of the 5,000 entries, 3,824 could be used as noun, verb, or both. The word class judgments were basically those taken from Hoffmann, but were further checked with one native speaker (Maki Purti). Because of the large number of items for this part of the study it was not feasible to determine directionality, or the behaviour of their English equivalents. The full list, contained in Osada (2004), is downloadable as a pdf file from http://munda.chikyu.ac.jp/EM/.

For the 3,824 eligible items in this sample, we find the figures given in Table 1. These figures give a substantially lower percentage of fluid flexible lexemes than those for our earlier sample, presumably reflecting a combination of sampling bias in the smaller sample (through the inclusion of Hoffmann's and Bhat's showpiece examples) and a lowered likelihood of the more obscure vocabulary items – present only in the larger sample – of exhibiting zero conversion. They confirm quite clearly that the occurrence of zero conversion in Mundari is of a comparable order to English: it is common, but not available without limit, and there exist large numbers of both nouns and verbs that do not have other syntactic possibilities available to them.[32] Our discussion in Sec-

32. Our findings thus run counter to Bhat's (1997: 243–244)interpretation of the frequency of conversion or fluidity in the *Encyclopaedia Mundarica*: "Hoffmann considers Mundari words to be of great functional elasticity, having very vague signifying power. He establishes this elasticity of Mundari word, rather convincingly, in his twelve-volume [*sic*] Mundari Encyclopedia, IN WHICH ALMOST EVERY WORD IS SHOWN [emphasis ours], with copious examples, to be occurring as a substantive, adjective, and also as a transitive and in transitive verb."

tion 3.2 showed that the semantic interpretation of convertible words cannot be predicted by general rules, and that though there are a half-dozen typical semantic relationships between the argumental and the predicate uses of lexemes these cannot be predicted from the ontological class of the lexeme. And our discussion in Section 3.4.1 demonstrated that, if we hold ontological type constant to just animal, plant, or kin terms, we still cannot predict whether conversion will be available or not. The percentage figures from this extended figures emphasize that non-convertibility is not a sporadic or marginal phenomenon, and that around half of the lexical items in our large sample do not have attested fluid uses. Of course it is always possible that our source fails to record some such uses, but it must be remembered that, with its sixteen substantial volumes, the Hoffmann *Encyclopaedia* constitutes one of the most exhaustive lexical documentations of any non-metropolitan language in the world.

4. Conclusion

The question of whether languages exist that lack a noun-verb distinction is a fundamental one for typology and for linguistics more generally, since much of the system of morphosyntactic rules is built on the generalizations holding at the level of word class. Since languages without word classes are clearly imaginable – Predicate Calculus being the canonical example – any finding that languages universally distinguish nouns from verbs represents a major constraint on the form of possible human languages (cf. Baker 2001). Though a number of claimed cases of such languages have been put forward, with some typological accounts (e.g., Hengeveld 1992a, b, Rijkhoff 2002) taking it as established that such languages exist, and though it is clear that in many languages there is only a "weak" noun-verb distinction, we do not believe there exist – as yet – attested cases of languages lacking a noun-verb distinction altogether, according to the highest standards of description and argumentation.

Because the theoretical stakes are so high, linguists must apply the most rigorous standards of proof to claimed cases, and part of our job as typologists is to establish standardized criteria of argumentation that can be used across all languages in our purview, at the same time establishing a typology of the different ways in which languages come close – or appear to come close – to lacking this distinction.

Our goal in this article has been to re-evaluate one such claimed case, viz. Mundari, in the light of both fresh data, and of more explicit lines of argumentation. Our verdict is that Munda clearly distinguishes nouns from verbs, though (like English, Chinese, and many other languages) it has widespread zero conversion, extending to around 50 % of the lexicon. In fact this finding agrees with Hoffmann's later view of the language once the compilation of a complete dictionary forced him to extend his analysis to the full lexicon.

At the same time, we have sought to spell out three general requirements that must be met before a language can be said to have a single merged class: DISTRIBUTIONAL EQUIVALENCE that is FULLY BIDIRECTIONAL, explicit SEMANTIC COMPOSITIONALITY for argument and predicate uses, and EXHAUSTIVENESS in the form of a demonstration that these effects hold over the complete lexicon, not just for a few favoured cases. It will be interesting to see how well other claimed "fluid" languages stand up to these tests.

Received: 8 September 2004
Revised: 24 December 2004

University of Melbourne
Research Institute for Humanity and Nature, Kyoto

Correspondence addresses: (Evans) Linguistics & Applied Linguistics, University of Melbourne, Parkville, Victoria 3010, Australia; e-mail: nrde@unimelb.edu.au; (Osada) Research Institute for Humanity and Nature, 335 Takashima-cho, Marutamachi-dori Kawaramachi nishi-iru, Kamigyo-ku, Kyoto 602-0878, Japan; e-mail: osada@chikyu.ac.jp

Acknowledgements: Our collaboration on this paper was made possible through the University of Melbourne's Young Asian Scholars scheme, which enabled Osada to spend time in Melbourne, with further assistance from a School of Languages Grant-in-Aid, from a Grant-in-Aid for Scientific Research to Osada from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (Grant (C) No. 13610677, years 2001–2003), which supported the database of entries from the *Encyclopaedia Mundarica*, and as a sideline from joint work on reciprocal constructions in Mundari supported by a grant to Evans from the Australian Research Council (Grant: Reciprocals Across Languages). We thank these bodies for their generous support. Maki Purti has assisted us at every stage with her expert native speaker judgments on Mundari. We further thank Bob Dixon, Sasha Aikhenvald, and two anonymous *LT* referees for their comments on earlier versions of this paper, including an oral presentation in June 2002 at the Research Centre for Linguistic Typology, La Trobe University.

Abbreviations: 1, 2, 3 1st, 2nd, 3rd person, ABS absolutive, AN animate, ASP aspect, AT actor topic, COMPL completive aspect, CON continuous, COP copula, DAT dative, DEF definite, DEM demonstrative, DET determiner, DU dual in suffixes, du dual in enclitics, EXCL exclusive, FEM feminine, FUT future, GEN genitive, IMPF imperfective, INAN inanimate, INCEP inceptive, INDIC indicative, INIT.PROG initiated progressive, INTR intransitive, IPF imperfective, LOC locative, LOC.COP locational copula, NEG negative, NOM nominative, O object, PASS passive, PERF perfective, PL plural in suffixes, pl plural in enclitics, PRES present, PROGR.OR progressive oriented, REL relative, S subject, STAT stative, TOP topic, TR transitive.

The nominal number suffixes in Mundari are written in small caps (DU, PL) to distinguish them from the (person plus) number enclitics and verbal suffixes (du, pl) with which they are homophonous.

Appendix: Lexical items included in our sample with summary of conversion possibilities

V → N		Eng. conv.	*V → N		Eng. conv.
rain	*gama*	+	cook	*isin*[a]	+
wind	*hoyo*	−	cut (with knife)	*had*	+
laugh	*landa*	+	sow	*her*	−
run	*nir*	+	make	*bai*	+
hit	*dal*	+	hear, listen	*ayum*	−
kick	*pada*	+	smell, sniff	*jii*	+
quarrel	*eperang*	+	fall over	*baTi*	+
fight	*gopoeq*	+	be able, can	*daRi*	−
fuck	*de(pe)Re*[b]	+	sit	*dub*	−
weave	*teng*	+	stand	*birid*	+
eat	*jom*	−	dawn	*ang*	−
drink	*nu*	+	bring	*au*	−
talk	*jagar*	+	flow	*atu*	+
give	*om*	+	take	*idi*	+
teach	*itu*	−			
see	*lel*	−			
rise	*rakab*	+			
die/kill	*goeq*	+(kill), −(die)			
cut with scissor-like motion	*laTab*	+			
*lie (posture)	*gitiq*	+			
(be) drunk	*bul* n. 'intoxication', v.t. 'to intoxicate', v.i. 'be drunk'	+			
divide, distribute, part	*haTing*	+			
evening	*ayub*	−			
go	*sen*	+			
come	*hijuq*	+			
catch	*sab*	+			
throw	*ter*	+			

v/n. (basic meaning and combinatorics difficult to decide)
rimbil 'cloud/cloudy', *durang* 'sing/song', *soan* 'stink, smell'

a Though there is a collocation with this verb which does allow conversion, namely *isin basang* 'boil. (v.t.); n. the water in which the rice is to be cooked'; *basang* = cooking.
b Hoffman does not give the nominal use in his *Encyclopaedia Mundarica* but Maki Purti assures us it is common.

N → V		Eng. conv.	N −/ → V		Eng. conv.
sun	*singgi*	+	man; male	*koRa*	+
fever	*rua*	−	father	*apu*	+
steal	*kumbuRu*	+	nose	*muu*	+
disease, sickness	*rogo*	−	weaver	*peNae*	−
old, old man	*haRam*	−	(place name)	*Ranci*	−
boy	*danggRa*	−	council, panchayat or other meeting	*saba*	+
girl	*danggRi*	−			
man; person	*hoRo*	+	forest	*bir*	+
wife, woman, girl	*kuRi*	−	dog	*seta*	+
stone	*diri*	+	pig	*sukuri*	−
dust	*duRa*	+	goat	*merom*	−
tree	*daRu*	−	cat	*pusi*	−
mother	*engga*	+	cattle	*uriq*	−
child	*hon*	−	fig	*loa*	+
brother	*haga*	−	mango	*uli*	−
mother-in-law	*hanar*	−	jackfruit	*kaNTaRa*	−
daughter-in-law	*kimin*	−	rice (paddy)	*baba* n.	−
head	*booq*	+	fish	*hai*	+
ear	*lutur*	−			
foot, leg	*kaTa*	+			
eye	*med*	+			
stomach	*laiq*	+			
hammer	*koTasi*	+			
bow and arrow	*aqsar*	+			
axe	*kapi*	+			
knife	*katu*	+			
pancake	*lad*	−			
cowherd	*mahara*	−			
smith, blacksmith	*baRae*	−			
night	*nida*	−			
mountain	*buru*	−			
shade	*umbul*	+			
river	*gaRa*	−			
village	*hatu*	−			
clan	*kili*	−			
tribe	*jati*	−			
custom	*dastur*	−			
fowl, chicken	*sim*	+			
tiger	*kula*	−			
flower	*baa*	−			

N → V		Eng. conv.	N −/→ V	Eng. conv.
tamarind (and pulp); sour; make sour	*jojo*	−		
food grain, rice	*maNDi*	−		
water	*daq*	+		
meat	*jilu*	−		
outsider, Hindu	*diku*	−		
Sunday	*etwar*	−		
sowing-time drinking feast	*herpuna*	+[a]		

a 'Christmas' is taken as a comparable English word here.

References

Adelaar, K. Alexander & Nikolaus Himmelmann (eds.) (2004). *The Austronesian Languages of Asia and Madagascar*. London: Routledge Curzon.

Anderson, Gregory D. S. & Norman Zide (2001). Recent advances in the reconstruction of the proto-Munda verb. In Laurel Brinton (ed.), *Historical Linguistics 1999*, 13–30. Amsterdam: Benjamins.

Anderson, Steven (1985). Inflectional morphology. In Timothy Shopen (ed.), *Language Typology and Linguistic Description*, Volume 3: *Grammatical Categories and the Lexicon*, 150–201. Cambridge: Cambridge University Press.

Andrews, J. Richard (1975). *Introduction to Classical Nahuatl.* Austin: University of Texas Press.

Baker, Mark (2001). *The Atoms of Language: The Mind's Hidden Rules of Grammar.* New York: Basic Books.

— (2003). *Lexical Categories: Verbs, Nouns and Adjectives.* Cambridge: Cambridge University Press.

Bhat, D. N. S. (1994). *The Adjectival Category: Criteria for Differentiation and Identification.* Amsterdam: Benjamins.

— (1997). Noun-verb distinction in Munda languages. In Anvita Abbi (ed.), *Languages of Tribal and Indigenous Peoples of India: The Ethnic Space*, 227–251. Delhi: Motilal Banarsidas.

— (2000). Word classes and sentential function. In Vogel & Comrie (eds.) 2000, 47–64.

Bodding, P. O. (1929). *Materials for a Santali grammar, II. Mostly Morphology*. Dumka: Santal Mission of the Northern Churches.

Broschart, Jürgen (1997). Why Tongan does it differently. *Linguistic Typology* 1: 123–165.

Churchward, C. Maxwell (1953). *Tongan Grammar.* London: Oxford University Press.

Clark, Eve V. & H. Herbert Clark (1979). When nouns surface as verbs. *Language* 55: 767–811.

Croft, William (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information.* Chicago: University of Chicago Press.

— (2000). Parts of speech as typological universals and language particular categories. In Vogel & Comrie (eds.) 2000, 65–102.

— (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

— (2003). *Typology and Universals*. 2nd edition. Cambridge: Cambridge University Press.

Demirdache, Hamida & Lisa Matthewson (1995). On the universality of syntactic categories. *North Eastern Linguistic Society* 25: 79–93.
Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language* 1: 19–80.
— (1988). *A Grammar of Boumaa Fijian.* Chicago: University of Chicago Press.
Dixon, R. M. W. & Alexandra Y. Aikhenvald (eds.) (2004). *Adjective Classes: A Cross-linguistic Typology*. Oxford: Oxford University Press.
Enfield, Nick (2004). Adjectives in Lao. In Dixon & Aikhenvald (eds.) 2004, 323–347.
Evans, Nicholas (2000a). Word classes in the world's languages. In Geert Booij, Christian Lehmann, & Joachim Mugdan (eds.), *Morphology: A Handbook on Inflection and Word Formation*, 708–732. Berlin: Mouton de Gruyter.
— (2000b). Kinship verbs. In Vogel & Comrie (eds.) 2000, 103–172.
Foley, William A. (1998). Symmetrical voice systems and precategoriality in Philippine languages. Paper presented at the 3rd LFG conference, Brisbane.
Gil, David (1994). The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17: 179–200.
— (1995). Parts of speech in Tagalog. In Mark Alves (ed.), *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society*, 67–90. Tempe, Arizona.
— (2001). Escaping Eurocentrism: Fieldwork as a process of unlearning. In Paul Newman & Martha Ratliff (eds.), *Linguistic Fieldwork*, 102–132. Cambridge: Cambridge University Press.
Hengeveld, Kees (1992a). Parts of speech. In Michael Fortescue, P. Harder, & L. Kristofferson (eds.), *Layered Structure and Reference in a Functional Perspective*, 29–53. Amsterdam: Benjamins.
— (1992b). *Non-verbal predication: Theory, Typology, Diachrony*. Berlin: Mouton de Gruyter.
Himmelmann, Nikolaus (1991). The Philippine challenge to Universal Grammar. *Arbeitspapiere des Instituts für Sprachwissenschaft der Universität zu Köln*, Neue Folge 15.
— (2004a). The Austronesian languages of Asia and Madagascar: Typological characteristics. In Adelaar & Himmelmann (eds.) 2004, 110–181.
— (2004b). Tagalog. In Adelaar & Himmelmann (eds.) 2004, 350–376.
Hoffmann, John (1903). *Mundari Grammar.* Calcutta: The Secretariat Press.
Hoffmann, John & Arthur van Emelen (1930–1979). *Encyclopaedia Mundarica.* 16 volumes. Patna: Government Printing.
Hopper, Paul J. & Sandra A. Thompson (1984). The discourse basis for lexical categories in Universal Grammar. *Language* 60: 703–752.
Jacobson, William (1979). Noun and verb in Nootkan. *British Columbia Provincial Museum Heritage Records* 4: 83–155.
Jelinek, Eloise (1995). Quantification in Straits Salish. In Emmon Bach, Eloise Jelinek, Angelika Kratzer, & Barbara H. Partee (eds.), *Quantification in Natural Languages*, 487–540. Dordrecht: Kluwer.
Jelinek, Eloise, & Richard A. Demers (1994). Predicates and pronominal arguments in Straits Salish. *Language* 70: 697–736.
Kinkade, M. Dale (1983). Salish evidence against the universality of "noun" and "verb". *Lingua* 60: 25–40.
Kroeger, Paul (1998). Nouns and verbs in Tagalog: A reply to Foley. Paper presented at the 3rd LFG conference, Brisbane.
Kuipers, Aert H. (1968). The categories verb-noun and transitive-intransitive in English and Squamish. *Lingua* 21: 610–626.
Langacker, Ronald W. (1987). Nouns and verbs. *Language* 63 (1): 53–94.
Langendoen, Terence D. (1967). The copula in Mundari. In John W. M. Verhaar (ed.), *The Verb 'Be' and its Synonyms*, 75–100. Dordrecht: Reidel.
Launey, Michel (1994). *Une grammaire omniprédicative.* Paris: CNRS Editions.
— (2002). Compound nouns vs. incorporation in Classical Nahuatl. In Nicholas Evans & Hans-Jürgen Sasse (eds.), *Problems in Polysynthesis*, 113–134. Berlin: Akademie-Verlag.

Lichtenberk, Frantisek (1991). Semantic change and heterosemy in grammaticalization. *Language* 67: 474–509.
Milner, George B. (1972) *Fijian Grammar.* 3rd edition. Suva: Government Press.
Mosel, Ulrike & Einar Hovdhaugen (1992). *Samoan Reference Grammar.* Oslo: Scandinavian University Press.
Munda, Ram Dayal (1971). Aspect of Mundari verb. *Indian Linguistics* 32: 27–49.
Neukom, Lukas (2001). *Santali.* München: Lincom Europa.
Norman, Jerry (1988). *Chinese*. Cambridge: Cambridge University Press.
Osada, Toshiki (1992). *A Reference Grammar of Mundari.* Tokyo: Institute for the Languages and Cultures of Asia and Africa.
— (1999). Experiential constructions in Mundari. *Gengokenkyu (Journal of the Japanese Linguistic Society)* 115: 51–76.
— (2004). The reexamination on noun/verb distinction in Mundari appeared in the selected entries of Encyclopaedia Mundarica. Unpublished manuscript, Research Institute for Humanity and Nature.
— (forthcoming). Reciprocals in Mundari. In Vladimir Nedjalkov & Zlata Guentcheva (eds.), *Typology of Reciprocal Constructions*. München: Lincom Europa.
Pinnow, Heinz-Jürgen (1960). Über den Ursprung der voneinander abweichenden Strukturen der Munda- und Khmer-Nikobar-Sprachen. *Indo-Iranian Journal* 4: 81–103.
— (1966). A comparative study of the verb in the Munda languages. In Norman Zide (ed.), *Studies in Comparative Austroasiatic Linguistics*, 96–193. The Hague: Mouton.
Ponette, P. (1990). An Historical Note on the *Encyclopaedia Mundarica*. Pp. i–xii of Hoffmann & Emelen, Vol. XV.
Rijkhoff, Jan (2002). *The Noun Phrase.* Oxford: Oxford University Press.
— (2003). When can a language have nouns and verbs? *Acta Linguistica Hafniensia* 35: 6–38.
Sasse, Hans-Jürgen (1993). Syntactic categories and subcategories. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, & Theo Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, 646–687. Berlin: de Gruyter.
Schachter, Paul (1985). Parts-of-speech systems. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, Volume 1: *Clause Structure*, 3–61. Cambridge: Cambridge University Press.
Schiller, Eric (1989). Syntactic polysemy and underspecification in the lexicon. *Berkeley Linguistics Society* 15: 278–290.
— (1992). Parts of speech in Southeast Asian languages: An autolexical view. In *Pan-Asiatic Linguistics: Proceedings of the 3rd International Symposium on Language and Linguistics, Vol. II*, 778–790.
Sinha, N. K. (1975). *Mundari Grammar.* Mysore: Central Institute of Indian Languages.
Stassen, Leon (1997). *Intransitive Predication.* Oxford: Oxford University Press.
Swadesh, Morris (1939). Nootka internal syntax. *International Journal of American Linguistics* 9: 77–102.
Thompson, Sandra (1988). A discourse approach to the cross-linguistic category 'adjective'. In John A. Hawkins (ed.), *Explaining Language Universals*, 167–208. Oxford: Blackwell.
Verhaar, John W. M. (1984). Affixation in contemporary Indonesian. In B. Kaswanti Purwo (ed.), *Towards a Description of Contemporary Indonesian: Preliminary Studies*, Part I (NUSA, 18), 1–26. Jakarta: Universitas Atma Jaya.
Vogel, Petra M. & Bernard Comrie (eds.) (2000). *Approaches to the Typology of Word Classes*. Berlin: Mouton de Gruyter.
Wetzer, Harrie (1996). *The Typology of Adjectival Predication.* Berlin: Mouton de Gruyter.
Whorf, Benjamin Lee (1956). Science and Linguistics. In J. B. Carroll (ed.), *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*, 207–219. New York: MIT Press.